Exercise 9 Joins in hive \

-- Create tables and load data

CREATE TABLE students (

student\_idINT,

student\_name STRING,

course STRING,

grade STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY','

STORED AS TEXTFILE;

CREATE TABLE courses (

course\_name STRING,

instructor STRING,

creditsINT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY','

STORED AS TEXTFILE;

Class: II M.Sc DS Semester: III

P23DS3P6: Big Data Management and Analytics Lab P a g e |38

-- Load data into the tables

LOAD DATA INPATH 'students.csv'INTOTABLE students;

LOAD DATA INPATH ‘courses.csv'INTOTABLE courses;

-- Inner Join: Get student names and course details for students enrolled in each

course

SELECTs.student\_name, c.course\_name, c.instructorFROM students s JOIN courses c

ONs.course=c.course\_name;

-- Left Join: Get all students and their corresponding course details, including

students not enrolled in any course

SELECTs.student\_name, c.course\_name, c.instructorFROMstudents sLEFTJOIN courses c

ONs.course=c.course\_name;

-- Aggregation: Calculate the average grade per course

SELECT course, AVG(

CASE grade

WHEN'A'THEN4

WHEN'B'THEN3

WHEN'C'THEN2

WHEN'D'THEN1

ELSE0

END

) ASaverage\_grade

FROM students

GROUPBY course;

-- Aggregation: Count the number of students enrolled in each course

SELECT course, COUNT(student\_id) ASstudent\_countFROM students GROUPBY course;

Sample Dataset:

 students.csv: {student\_id,student\_name,course,grade}

1,JohnDoe,Mathematics,A

2,JaneSmith,Science,B

3,EmilyDavis,Mathematics,C

4,MichaelBrown,History,A

 courses.csv: {course\_name,instructor,credits}

Mathematics,Dr. Smith,3

Science,Dr. Johnson,4

History,Dr. Lee,3

**Start Hive CLI:**

bash

Copy code

hive

**Create a Database (if needed):**

sql

Copy code

CREATE DATABASE IF NOT EXISTS company\_db;

**Use the Created Database:**

sql

Copy code

USE company\_db;

**Create the Employees Table:**

sql

Copy code

CREATE TABLE IF NOT EXISTS employees (

employee\_id INT,

employee\_name STRING,

department\_id INT,

salary FLOAT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

**Create the Departments Table:**

sql

Copy code

CREATE TABLE IF NOT EXISTS departments (

department\_id INT,

department\_name STRING

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

STORED AS TEXTFILE;

**Load Data into the Employees Table:**

sql

Copy code

LOAD DATA LOCAL INPATH '/path/to/employees\_data.csv'

INTO TABLE employees;

**Sample employees\_data.csv:**

mathematica

Copy code

1,John Doe,1,55000

2,Jane Smith,2,80000

3,Robert Brown,3,60000

4,Emily Davis,1,58000

5,Michael Johnson,2,90000

**Load Data into the Departments Table:**

sql

Copy code

LOAD DATA LOCAL INPATH '/path/to/departments\_data.csv'

INTO TABLE departments;

**Sample departments\_data.csv:**

Copy code

1,Sales

2,IT

3,HR

**Perform a Join between Employees and Departments:**

sql

Copy code

SELECT e.employee\_id, e.employee\_name, d.department\_name, e.salary

FROM employees e

JOIN departments d

ON e.department\_id = d.department\_id;

**Calculate Average Salary by Department:**

sql

Copy code

SELECT d.department\_name, AVG(e.salary) AS avg\_salary

FROM employees e

JOIN departments d

ON e.department\_id = d.department\_id

GROUP BY d.department\_name;

**Get Total Employees per Department:**

sql

Copy code

SELECT d.department\_name, COUNT(e.employee\_id) AS total\_employees

FROM employees e

JOIN departments d

ON e.department\_id = d.department\_id

GROUP BY d.department\_name;